Claims

What is claimed is:

1. A method of adapting a speech recognition system to speech data provided to the speech recognition system, the method comprising the steps of:

computing alignment information between the speech recognition system and feature vectors associated with the speech data provided to the speech recognition system;

computing an original spectra for each feature vector and corresponding mean vector;

estimating one or more transformation parameters which maximize a likelihood of an utterance; and

transforming a current feature vector using the estimated transformation parameters and maximum likelihood criteria, the transformation being performed in a linear spectral domain.

2. The method of claim 1, wherein the step of transforming the current feature vector is performed in feature space.

3. The method of claim 1, wherein the step of transforming the current feature vector is performed in model space.

4. The method of claim 1, wherein the maximum likelihood criteria is a maximum likelihood spectral transformation (MLST).

5. The method of claim 1, wherein the step of estimating one or more transformation parameters which maximize a likelihood of an utterance further comprises

the step of computing likelihood of utterance information corresponding to a previous feature vector transformation.

6. The method of claim 1, wherein the step of estimating the transformation parameters further comprises the step of estimating convolutional noise N_i^{α} and additive noise N_i^{β} for each ith component of a speech vector corresponding to the speech data provided to the speech recognition system.

7. The method of claim 6, wherein the step of estimating the transformation parameters further comprises the step of defining a diagonal matrix A with $A_{ii} = 1/N_i^{\alpha}$, and defining $b_i = -N_i^{\beta}/N_i^{\alpha}$.

8. The method of claim 7, further comprising the steps of:

determining A_{ii} in accordance with an expression

$$A_{ii} = \frac{T\sum_t x_{t,i}^{(\varepsilon)} m_{t,i}^{(\varepsilon)} - \sum_t x_{t,i}^{(\varepsilon)} \sum_t m_{t,i}^{(\varepsilon)}}{T\sum_t x_{t,i}^{(\varepsilon)2} - \sum_t x_{t,i}^{(\varepsilon)} \sum_t x_{t,i}^{(\varepsilon)}}\text{ ; and}$$

determining b_i in accordance with an expression

$$b_i = \frac{-A_{ii} \sum_t x_{t,i}^{(\varepsilon)} + \sum_t m_{t,i}^{(\varepsilon)}}{T};$$

where $x_{t,i}^{(\varepsilon)}$ and $m_{t,i}^{(\varepsilon)}$ are sub-linear spectral values of a feature vector and corresponding mean vector, respectively, for each ith component of the speech vector.

9. The method of claim 1, wherein the step of transforming the current feature vector further comprises the step of determining $\dot{x}_i^{(f)} = \frac{1}{N_i^{\alpha}} x_i^{(f)} - \frac{N_i^{\beta}}{N_i^{\alpha}}$, where $x_i^{(f)}$ is an ith component of a speech vector corresponding to the speech data provided to the speech recognition system, N_i^{α} is convolutional noise and N_i^{β} is additive noise of the ith component of the speech vector.

10. The method of claim 1, wherein the step of computing alignment information is performed using a Baum-Welch algorithm.

11. In a speech recognition system, a method of transforming speech feature vectors associated with speech data provided to the speech recognition system, the method comprising the steps of:

receiving likelihood of utterance information corresponding to a previous feature vector transformation;

estimating one or more transformation parameters based, at least in part, on the likelihood of utterance information corresponding to a previous feature vector transformation; and

transforming a current feature vector based on at least one of maximum likelihood criteria and the estimated transformation parameters, the transformation being performed in a linear spectral domain.

12. The method of claim 11, wherein the step of transforming the current feature vector is performed in feature space.

13. The method of claim 11, wherein the step of transforming the current feature vector is performed in model space.

14. The method of claim 11, wherein the maximum likelihood criteria is a maximum likelihood spectral transformation (MLST).

15. Apparatus for adapting a speech recognition system to speech data provided to the speech recognition system, the apparatus comprising:

at least one processing device operative to: (i) compute alignment information between the speech recognition system and feature vectors associated with the speech data provided to the speech recognition system; (ii) compute an original spectra for each feature vector and a corresponding mean vector; (iii) estimate one or more transformation parameters which maximize a likelihood of an utterance; and (iv) transform a current feature vector based on at least one of maximum likelihood criteria and the estimated transformation parameters, the transformation being performed in a linear spectral domain.

16. The apparatus of claim 15, wherein the operation of transforming the current feature vector is performed in a feature space.

17. The apparatus of claim 15, wherein the operation of transforming the current feature vector is performed in a model space.

18. The apparatus of claim 15, wherein the spectral transformation employed in the operation of transforming the current feature vector is a maximum likelihood spectral transformation (MLST).

19. The apparatus of claim 15, wherein the operation of estimating one or more transformation parameters which maximize a likelihood of an utterance further comprises the operation of computing likelihood of utterance information corresponding to a previous feature vector transformation.

20. The apparatus of claim 15, wherein the operation of estimating the transformation parameters further includes the operation of estimating convolutional

noise N_i^α and additive noise N_i^β for each ith component of a speech vector provided to the speech recognition system.

21. The apparatus of claim 20, wherein the operation of estimating the transformation parameters further includes the operation of defining a diagonal matrix A with $A_{ii} = 1/N_i^\alpha$, and defining $b_i = -N_i^\beta/N_i^\alpha$.

22. The apparatus of claim 21, wherein the operation of estimating the transformation parameters further comprises the operation of:

determining A_{ii} in accordance with an expression

$$A_{ii} = \frac{T\sum_t x_{t,i}^{(\varepsilon)} m_{t,i}^{(\varepsilon)} - \sum_t x_{t,i}^{(\varepsilon)} \sum_t m_{t,i}^{(\varepsilon)}}{T\sum_t x_{t,i}^{(\varepsilon)2} - \sum_t x_{t,i}^{(\varepsilon)} \sum_t x_{t,i}^{(\varepsilon)}} \text{ ; and}$$

determining b_i in accordance with an expression

$$b_i = \frac{-A_{ii} \sum_t x_{t,i}^{(\varepsilon)} + \sum_t m_{t,i}^{(\varepsilon)}}{T};$$

where $x_{t,i}^{(\varepsilon)}$ and $m_{t,i}^{(\varepsilon)}$ are sub-linear spectral values of a feature vector and corresponding mean vector, respectively, for each ith component of the speech vector.

23. The apparatus of claim 15, wherein the operation of transforming the current feature vector includes the step of determining $\dot{x}_i^{(f)} = \frac{1}{N_i^\alpha} x_i^{(f)} - \frac{N_i^\beta}{N_i^\alpha}$, where $x_i^{(f)}$ is an ith component of a speech vector corresponding to the speech data provided to the speech recognition system, N_i^α is convolutional noise and N_i^β is additive noise of the ith component of the speech vector.